

Implementation of local quality improvement collaboratives to improve test ordering and prescribing performance of general practitioners

Citation for published version (APA):

Trietsch, J. (2016). *Implementation of local quality improvement collaboratives to improve test ordering and prescribing performance of general practitioners*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20161021jt>

Document status and date:

Published: 01/01/2016

DOI:

[10.26481/dis.20161021jt](https://doi.org/10.26481/dis.20161021jt)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

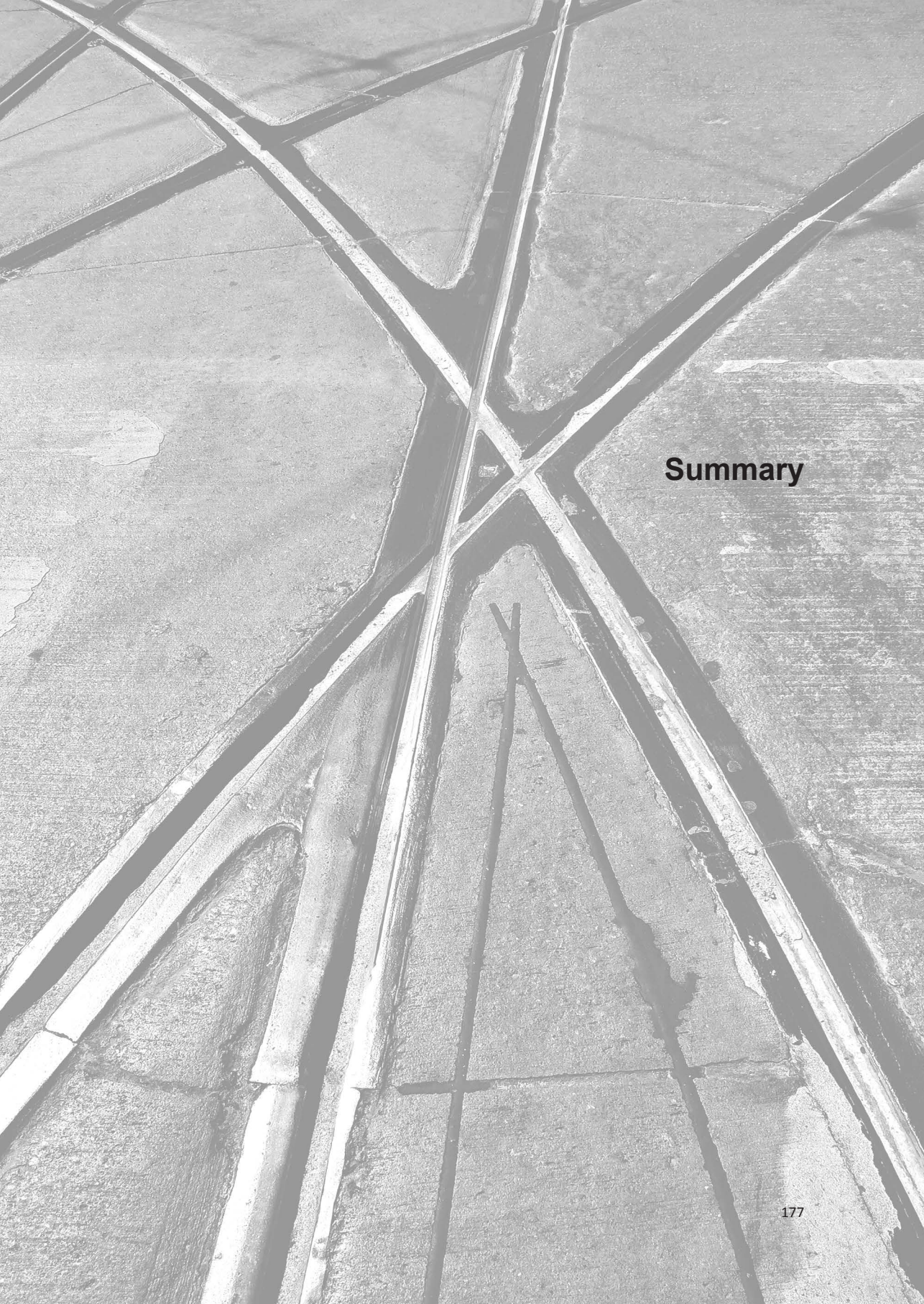
Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 05 May. 2023



Summary

Summary

This thesis describes the research that was put into a complex intervention aimed at general practitioners with the aim to improve their test ordering behavior and prescribing behavior. Interventions of this type were performed before but never on test ordering and prescribing together. Neither was this type of intervention ever tested under normal practice conditions with much freedom for the participants to adapt the intervention to their own needs and situations. In this chapter the separate sections of this thesis are summarized for a quick overview of the findings.

Chapter 1 Is the general introduction to this thesis. The background of the research is reported together with the questions this thesis addresses.

Several problems are present in healthcare today such as the ever growing costs of healthcare that policymakers need to control. Also physicians can't keep up with all the evidence that is being published. In the Netherlands therefore guidelines are being developed since the 1980's as a means to summarize evidence and translate it to the field. Researchers also started off to research how to effectively translate evidence to field workers. The most promising tools for this seem to be interventions that are multifaceted and contain education, audit and feedback, small peer group work and the use of opinion leaders. However, most of this research was conducted under well controlled circumstances which diminishes the external validity of these findings. Usually these interventions also dealt with only one or two well described clinical topics. As guidelines became available to general practitioners (GPs) they started to discuss prescribing behavior in small peer review groups together with their affiliated local pharmacist, so called pharmacotherapeutic audit groups (PTAM).

In an earlier trial small peer group groups were used for discussing test ordering behavior of GPs. Experiences from this intervention and the availability of an existing infrastructure of PTAM groups led us to designing a complex intervention to support GPs in following guidelines on test ordering and prescribing. The aim of this intervention was to transfer the strategy to the field.

Qualitative research was used to evaluate the transference of the intervention to the field. A specific design was used to research the effect on test-ordering rates and prescribing rates of this intervention. This design is known as a balanced incomplete block design in implementation science.

We searched for answers to the following questions in this thesis on changing test ordering behaviour and prescribing behaviour of GPs:

- Was the intervention executed as planned?
- What were the barriers to and facilitators of the implementation of the strategy?

- Has the level of group performance been improved in the participating groups?
- Did the volumes of tests ordered and drugs prescribed change in the preferred direction, as described in the working agreements of the local quality improvement collaboratives (LQICs), compared to baseline?
- What is the effect of this strategy on GPs' test ordering and prescribing behaviour in terms of interphysician variation and total volumes of tests and prescriptions with respect to specific clinical topics, compared to that among GPs exposed to the same strategy but for other topics?
- Is the balanced incomplete block design suitable for use in implementation science?

Chapter 2 is the research protocol for the intervention. In this chapter earlier work of our research group on test-ordering behavior is described. The results of the previous trials and the design of the present intervention are described in more detail. In the intervention comparative feedback on test-ordering and prescribing behavior of GPs was provided embedded in existing PTAM groups. In these PTAM groups the feedback will be discussed in a peer review process and working agreements will be set. The feedback will be offered during two consecutive sessions; the first on test ordering and the second on prescribing. The feedback offers standardized data on practice level together with data at group level and of the surrounding groups. Hereby these PTAMs will change their names to local quality improvement collaboratives.

The methodology of the study is described in detail. It is a cluster randomized controlled trial (C-RCT) with two arms where both arms receive the same intervention but on different clinical topics. Both arms serve as blind controls for the other research arm. The main outcome measures were the change in volumes of tests ordered and volumes of drugs prescribed per 1000 patients per six months. Additionally the working agreements set by the GPs were analyzed and related to changes in volumes that were observed.

Chapter 3 reports the results of a process evaluation of the quality improvement strategy consisting of comparative feedback and peer review embedded in already existing local quality improvement collaboratives (research questions 1 and 2). All elements of the intervention were of proven effectiveness and feasibility on changing physician behavior in previous trials. However, implementing our combined strategy covering both test ordering and prescribing performance proved to be problematic. The aim of this study was to describe and analyze the challenges of the transferring process. We interviewed 19 regional health officers, pharmacists, laboratory specialists and general practitioners within 6 months after the transfer period. The transcripts of these

interviews were independently coded. These codes were then matched to the dimensions of the normalization process model (NPM). This model can be used to structure factors that matter when normalizing an intervention. Normalizing can be interpreted here as the sustainable implementation of an intervention in a way that the groups will accept it and adopt it as normal.

The result of the evaluation showed that the general idea of the intervention was widely supported by GPs. The process of designing and maintaining the databases containing test ordering data and prescribing data needed to generate feedback as well as the generation of the feedback from these databases was more complex than expected. Also the participants did not assume responsibility for the work and the distribution of the resources that came with it as was intended by us. This resulted in a higher need for support from the research group than anticipated on. The conclusion is that although we used elements for our intervention that were proven to be effective and embedded the intervention in an existing infrastructure of PTAMs this proved to be insufficient to normalize the intervention. The general positive attitude towards our intervention did not change this.

Several elements came up that require special attention when designing such a complex intervention in the future. How to build and manage such large databases, the transference of the responsibility for tasks and the distribution of resources are the main elements that should be taken care of as early as possible when planning such a complex quality improvement strategy. Only copying the barriers and facilitators experienced in a preceding trial is not sufficient. However, despite such a preparation management decisions could change the healthcare field, resulting in the intervention being less fitted to the situation. We learned that multifaceted implementation strategies to change professional behavior seem attractive but they have a pitfall in their inherent complexity.

Chapter 4 describes the design that was used before in similar implementation research and the problems that cling to the use of its terminology in quality of care research. The balanced incomplete block (BIB) design has oftentimes been claimed to be used in designing and evaluating complex interventions in quality of care research. We have doubts about the appropriateness of using this design for this type of interventions. This chapter describes that various articles on health services research claim to have applied the BIB. After careful reading the design used seems to differ from the descriptions of the BIBD in original methodological literature on this design. The design used resembles more of a straight two armed trial or a classical Latin square design. The BIBD is a sophisticated design with the aim to reduce the need for resources in a surrounding where these resources are limited. We conclude that the

use of this design is not suited for evaluating complex interventions in research using humans. Luckily the statistical techniques used in the evaluation of the studies that claimed to have used the BIBD are appropriate. They were in fact appropriate for the evaluation of normal two armed experiments. Therefore the results that were presented are still valid. The use of the term BIBD in quality of care research seems inappropriate. Also more attention should be paid to proper referencing to the original methodological literature.

Chapter 5 describes the results of our cluster-RCT in existing local quality improvement collaboratives (LQICs) in general practice. Twenty-one LQIC groups, including 197 GPs working in 88 practices, entered the trial and were allocated to one of two trial arms. The intervention in both arms comprised of audit and feedback combined with peer review. Each arm was presented a different set of five clinical topics. The collaboratives chose three of the five clinical topics presented to them, according to their preference, and agreed on targets for change. The results of both a crude pre-post analysis of the differences in testing and prescribing rates at the LQIC level, and a per-protocol analysis for the same differences are presented. We did not demonstrate a difference in the mean changes in test ordering or prescribing volumes between intervention and control groups for the crude analysis. It was observed that the LQICs that deviated most from the mean volumes at baseline, changed most in the desired direction on both test ordering and prescribing performance. Therefore we conclude that the beneficial results from earlier work could not be confirmed by our attempt to implement the strategy in the field. It might be interesting to further research on the groups or practices that deviate the most from the mean at baseline. Targeting them for interventions on quality of care could potentially yield the largest effects. It remains unclear however whether this behavior is consistent behaviour for a single practice or group.

Chapter 6 addresses an overview of the results provided in this thesis and the discussion based on these results. We conclude that despite the vigorous design, the use of an existing backbone of LQICs and much room for local tailoring we did not succeed in effectively changing behavior on test ordering or prescribing by GPs. The uptake of the strategy by the field was also not as expected. It showed that the burden of running the intervention and the administration that came was relatively high. The administrative burden perceived by the Dutch GPs was already high due to a large healthcare reform. Would the strategy be implemented in all LQIC groups nationwide the costs of running the intervention would be very high and probably exceed the savings. Also the barriers and facilitators that we expected and anticipated on proved

to be incorrect. As this intervention is one of the first where positive findings from earlier well controlled interventions are tested under normal more or less uncontrolled circumstances some questions remain and arise from this thesis. Further research is needed to test whether the results from implementation research in well controlled settings can be replicated when tested under usual circumstances. The effects that we observed in the practices deviating most from the mean at baseline are interesting. This raises the question whether these GPs should better targeted instead of the whole field. The cut-off point for the definition of deviation and whether this behavior is consistent within a practice requires specific research.



Samenvatting

Samenvatting

Dit proefschrift beschrijft de resultaten van een complexe implementatiestrategie die als doel heeft om het diagnostisch aanvraag- en voorschrijfgedrag van huisartsen te verbeteren. Interventies van dit type die eerder zijn onderzocht waren gericht op het veranderen van diagnostisch aanvraag- of voorschrijfgedrag van artsen, maar nooit op beide gedragingen tegelijk. Ook is dit type strategie nooit eerder getest in de dagelijkse zorg waarbij veel vrijheid werd gegeven aan de deelnemers om de strategie aan te passen aan hun eigen behoeften en omstandigheden. In dit hoofdstuk worden de verschillende onderdelen van dit proefschrift samengevat ten behoeve van een snel overzicht van de bevindingen van ons onderzoek.

Hoofdstuk 1 is de algemene inleiding van dit proefschrift en beschrijft de achtergrond van het onderzoek.

De hedendaagse gezondheidszorg kent verschillende problemen, zoals onder andere de toenemende kosten van de gezondheidszorg die de beleidsmakers proberen te beheersen. Daarnaast kunnen artsen nieuwe kennis en inzichten die worden gepubliceerd nauwelijks toepassen door de grote hoeveelheid ervan. In Nederland zijn daarom sinds 1980 richtlijnen ontwikkeld als een middel om bevindingen uit de wetenschap samen te vatten en te vertalen naar het veld. Onderzoekers zochten naar effectieve manieren om wetenschappelijke bevindingen over te brengen aan de artsen. Meest veelbelovend lijken gecombineerde interventies bestaande uit educatie, audit en feedback, intercollegiale toetsing in kleine groepen en het gebruik van rolmodellen. Het grootste deel van dergelijk onderzoek werd echter uitgevoerd onder goed gecontroleerde omstandigheden waardoor de externe validiteit van deze bevindingen relatief laag is. Ook behandelden deze interventies meestal slechts één of twee nauw afgebakende klinische onderwerpen. Tegelijkertijd met de ontwikkeling van richtlijnen begonnen huisartsen, samen met hun lokale apotheker, met het bespreken van hun voorschrijfgedrag in kleine groepen, zogenaamde farmacotherapeutische overleggen (FTO's). Uitgaande van de bestaande infrastructuur van FTO's en eerdere ervaringen met intercollegiale toetsing over diagnostiek, het zogenaamde Diagnostisch Toets Overleg (DTOs), hebben we een complexe implementatiestrategie ontworpen om huisartsen in het veld te ondersteunen in het navolgen van richtlijn-aanbevelingen over aanvraag- en voorschrijfgedrag. Hierbij hadden we de ambitie om de strategie over te dragen aan het veld zelf. De haalbaarheid van het uitvoeren van de strategie door het veld zelf hebben we geëvalueerd door middel van kwalitatief onderzoek. Het effect op het aanvraag- en voorschrijfgedrag hebben we geëvalueerd door middel van een

specifiek design, in de implementatie-literatuur bekend als het zogenaamde balanced incomplete block design (BIBD).

De vragen die we willen beantwoorden in dit proefschrift over diagnostisch aanvraag- en voorschrijfgedrag van huisartsen zijn:

1. Werd de implementatiestrategie uitgevoerd zoals gepland?
2. Wat waren de belemmeringen en de bevorderende factoren in de uitvoering van de strategie?
3. Is het niveau van presteren van de deelnemende FTO-DTO groepen verbeterd? Zijn de volumina van de aangevraagde tests en voorgeschreven medicijnen veranderd in de gewenste richting, zoals beschreven in de werkafspraken van de groepen zelf, dit in vergelijking tot de Ausgangssituatie?
4. Wat is het effect van deze strategie op het diagnostisch aanvraag- en voorschrijfgedrag van huisartsen in termen van interdokter variatie en het totale volume van diagnostische aanvragen en voorschriften van medicatie met betrekking tot specifieke klinische onderwerpen? Dit in vergelijking met huisartsen die dezelfde strategie aangeboden kregen maar over andere onderwerpen.
5. Is het BIBD geschikt voor gebruik in onderzoek naar implementatiestrategieën?

Hoofdstuk 2 beschrijft het onderzoeksprotocol voor de implementatiestrategie. In dit hoofdstuk wordt het eerdere werk van onze onderzoeksgroep ten aanzien van FTO en DTO beschreven. De resultaten uit deze eerdere studie en het ontwerp van de huidige interventie worden verder uitgewerkt. De implementatiestrategie bestaat uit het verstrekken van feedback met vergelijkende spiegelinformatie over aanvraag- en voorschrijf volumina van de participerende huisartsen. De spiegelinformatie wordt vervolgens besproken in het bestaande FTO (intercollegiale toetsing), waarin afspraken voor toekomstig beleid worden gemaakt. De spiegelinformatie wordt in twee sessies aangeboden over hetzelfde onderwerp. De eerste sessie gaat over het diagnostisch aanvraag gedrag en de tweede sessie gaat over het voorschrijven van medicatie. De spiegelinformatie is samengevat op praktijkniveau en wordt vergeleken met de cijfers van de groep en de omliggende huisartsgroepen.

Daarnaast wordt de methodologie van de studie beschreven. Deze studie betreft een cluster gerandomiseerd gecontroleerd experiment met twee onderzoekarmen waarbij alle huisarts-groepen in beide armen dezelfde verbeter-interventie zoals zojuist beschreven aangeboden kregen, maar per onderzoekarm over verschillende klinische onderwerpen. Beide armen dienden als geblindeerde controles voor de andere onderzoekarm. De belangrijkste uitkomstmaten waren de verandering in de volumina van aangevraagde tests en voorgeschreven geneesmiddelen per 1000 patiënten per half jaar. Daarnaast werden de werkafspraken die de huisartsen maakten geanalyseerd

en gerelateerd aan de veranderingen in de gestandaardiseerde volumina die werden waargenomen.

In **Hoofdstuk 3** worden de resultaten beschreven van de procesevaluatie van de implementatiestrategie, bestaande uit vergelijkende feedback en intercollegiale toetsing ingebed in de bestaande FTO groepen (onderzoeksvragen 1 en 2). Alle elementen van de strategie zijn in eerdere onderzoeken effectief en haalbaar gebleken in het veranderen van het gedrag van artsen. Onze uitvoering van de gecombineerde strategie die zowel diagnostisch aanvraag gedrag als het voorschrijfgedrag van huisartsen probeerde te veranderen bleek echter problematisch. Het doel van dit onderdeel van de studie was om de problemen in het proces van implementatie te beschrijven en te analyseren. We interviewden 19 regionale medisch managers in de eerste lijn, apothekers, laboratorium-specialisten en huisartsen binnen 6 maanden na afloop van de interventie. De transcripties van deze interviews werden onafhankelijk gecodeerd. Deze codes werden daarna ingepast in de dimensies van het normalization process model (NPM). Dit model kan worden gebruikt om factoren die van belang zijn om te komen tot normalisatie van een interventie te structureren. Normaliseren staat hierbij voor het effectief en langdurig implementeren van een strategie, zodanig dat deze door de groep als standaard (normaal) uitgevoerd wordt.

Uit het resultaat van de evaluatie bleek dat huisartsen de idee achter de interventie op grote schaal ondersteunden. Het opzetten en onderhouden van de voor de feedback benodigde databases met aanvraag- en voorschrijfgegevens en het genereren van de feedback hieruit bleek complexer dan vooraf verwacht. Daarnaast bleek dat de deelnemers niet de verantwoordelijkheid namen voor het werk en de verdeling van de middelen, zoals door ons bedoeld. Hierdoor was er een grotere behoefte aan ondersteuning door de onderzoeksgroep dan verwacht en bedoeld vooraf. Ondanks dat de elementen waaruit onze strategie opgebouwd was in eerder onderzoek effectief bleken en de strategie uitgevoerd werd in de bestaande infrastructuur van FTO groepen bleek dit onvoldoende om de interventie normaliseren. De algemene positieve houding ten opzichte van onze interventie heeft dit helaas niet kunnen veranderen.

Uit de resultaten van deze procesevaluatie komen verschillende punten naar voren die speciale aandacht nodig hebben bij het ontwerpen van dergelijke complexe implementatiestrategieën in de toekomst. Ten eerste dient vooraf veel aandacht gegeven te worden aan het opzetten en beheren van grote databases voor feedback. Ten tweede moet de overdracht van de verantwoordelijkheid voor taken en de verdeling van de middelen zo vroeg mogelijk worden overdacht bij het plannen van een dergelijke complexe strategie. Alleen het overnemen van de weerstanden en bevorderende factoren, zoals ervaren in eerdere onderzoeken, blijkt onvoldoende. Een

dergelijke voorbereiding kan echter niet voorkomen dat door onverwachte veranderingen in het management van de gezondheidszorg de strategie minder goed past dan voorzien. Wij hebben geleerd dat complexe strategieën om professioneel gedrag te veranderen aantrekkelijk lijken, maar dat deze complexiteit ook een valkuil is.

Hoofdstuk 4 beschrijft de onderzoeksopzet en de bezwaren die kleven aan de terminologie die ervoor wordt gehanteerd in de wetenschappelijke literatuur. Het balanced incomplete block design (BIBD) wordt vaak beweerd gebruikt te zijn bij het ontwerpen en evalueren van complexe interventies in kwaliteit van zorg onderzoek. Wij hebben twijfels over de geschiktheid van het gebruik van dit design voor dergelijke interventies. Dit hoofdstuk beschrijft dat verschillende artikelen over gezondheidszorgonderzoek beweren het BIBD gebruikt te hebben. Bij nalezen lijkt de gebruikte onderzoeksopzet te verschillen van de omschrijving van het BIBD in de originele methodologische literatuur over dit ontwerp. De gebruikte onderzoeksopzet lijkt eerder een normaal twee armig experiment of eventueel een Latin square opzet. Het BIBD is een geavanceerd en complex design met als doel om onderzoek mogelijk te maken in een omgeving waarin beperkte middelen beschikbaar zijn. We concluderen dat het gebruik van dit design niet geschikt is voor het evalueren van complexe interventies in mens gebonden onderzoek. Gelukkig zijn de statistische technieken die gebruikt zijn bij de evaluaties in de studies die beweerden het BIBD toegepast te hebben wel degelijk geschikt. Het betrof in feite statistische technieken voor de evaluatie van normale tweearmige experimenten. Derhalve zijn de resultaten die gepresenteerd zijn in deze onderzoeken nog steeds geldig. Het gebruik van de term BIBD in onderzoek gericht op verbetering van de kwaliteit van zorg lijkt oneigenlijk. Ook zou meer aandacht moeten worden besteed aan juiste verwijzingen naar originele methodologische literatuur.

Hoofdstuk 5 beschrijft de resultaten van ons cluster-experiment in bestaande FTO groepen in de huisartsgeneeskunde in Nederland. Eenentwintig FTO groepen, bestaande uit 197 huisartsen werkzaam in 88 praktijken, namen deel en zijn aan één van de twee onderzoekarmen toegewezen. De implementatiestrategie in beide armen bestond uit audit en feedback in combinatie met intercollegiale toetsing. De groepen in beide armen kregen echter een andere set van vijf klinische onderwerpen aangeboden. Iedere FTO groep koos drie van de vijf klinische onderwerpen, op basis van hun eigen voorkeur, en stelden doelen voor verandering van beleid binnen hun eigen praktijken voor. De resultaten van zowel de intention-to-treat voor-na analyse van de verschillen in volumina van aangevraagde diagnostiek en de voorgeschreven medicatie op praktijkniveau, en een per-protocol analyse van dezelfde verschillen worden in dit

hoofdstuk gepresenteerd. We hebben gemiddeld gezien geen verschil gevonden in de verandering van volumina van diagnostisch aanvraag- en voorschrijfgedrag tussen interventie- en controlegroep in de intention-to-treat analyse. Wel zagen wij dat de huisartsgroepen die het meest afweken van de gemiddelde volumina bij aanvang van de studie, de grootste verandering lieten zien in de gewenste richting. Op basis van deze resultaten concluderen wij dat de positieve resultaten uit eerder werk niet konden worden bevestigd na onze poging om de strategie te implementeren in het bestaande netwerk van FTO groepen. Het zou interessant zijn om verder onderzoek te doen naar effecten van verbeterinterventies uitsluitend gericht op praktijken of FTO groepen die het meest van het gemiddelde afwijken bij een nulmeting. Het is echter vooralsnog onduidelijk of dergelijk afwijkend gedrag van een praktijk op een nulmeting voorspellend is voor afwijkend gedrag op andere tijdstippen en op andere domeinen.

Hoofdstuk 6 bevat de discussie van de studieresultaten die beschreven zijn in dit proefschrift. We concluderen dat ondanks de het gebruik van een gecombineerde strategie, het gebruik van het bestaande netwerk van FTO groepen en veel ruimte voor lokaal maatwerk, we er niet in geslaagd zijn om diagnostisch aanvraag- en voorschrijfgedrag door huisartsen effectief te veranderen. Ook verliep de incorporatie van de strategie in de bestaande FTO structuur door het veld niet zoals beoogd. De administratieve last van het uitvoeren van de interventie was relatief hoog, zeker gezien de administratieve last die veel huisartsen de laatste jaren ervaren door de veranderingen in de Nederlandse gezondheidszorg (transparantie, declaraties). Zou deze strategie landelijk in alle FTO groepen uitgerold worden dan zouden de kosten in tijd en geld voor het uitvoeren van de strategie niet opwegen tegen de besparingen.

De belangrijkste redenen voor het falen van onze strategie zijn een algemeen gevoel van overbelast zijn van huisartsen als gevolg van een grote gezondheidszorg hervorming die kort na het begin van de werving werd geïnitieerd. Daarbovenop bleken wij een onjuiste of op zijn minst achterhaalde visie te hebben op de te verwachten belemmerende en bevorderende factoren. Aangezien deze interventie een van de eersten is waar positieve bevindingen uit eerder goed gecontroleerde evaluatiestudies worden getest onder normale, min of meer ongecontroleerde omstandigheden blijven een aantal vragen onbeantwoord en worden enkele nieuwe vragen opgeworpen. Verder onderzoek is nodig om te testen of de resultaten van implementatieonderzoek dat uitgevoerd is onder goed gecontroleerde omstandigheden kan worden gerepliceerd indien het getest wordt onder normale omstandigheden. De effecten die we waarnamen in de praktijken die het meest afweken van de gemiddelde uitgangswaarden zijn interessant. Het roept de vraag op of toekomstig onderzoek zich beter op deze huisartsen zou kunnen richten in plaats van op het hele veld. De

afkapwaarde voor de definitie van “afwijkend” en de vraag of dit gedrag consistent is binnen een praktijk vereist verder onderzoek.

.